## **About Dataset**

## **1. Data Source:**

Synthetic data generated from the Wharton Class of 2025's statistics.

## **2. Meta Data:**

* application\_id: Unique identifier for each application
* gender: Applicant's gender (e.g., Male, Female)
* international: International student (TRUE/FALSE)
* gpa: Grade Point Average of the applicant
* major: Undergraduate major (e.g., STEM, Humanities)
* race: Racial background of the applicant (e.g., White, Black, Asian, Other / null: international student)
* gmat: GMAT score of the applicant
* work\_exp: Number of years of work experience (Year)
* work\_industry: Industry of the applicant's previous work experience (e.g., Finance, Technology)
* admission: Admission status (e.g., Admit, Waitlist, Null: Deny)
* Age
* url

## **3. Usage:**

* Exploratory Data Analysis (EDA): Understand the distributions, relationships, and patterns within the data.
* Classification: Predict the admission status based on other features.

## **3. Usage:**

* Exploratory Data Analysis (EDA): Understand the distributions, relationships, and patterns within the data.
* Classification: Predict the admission status based on other features.

## **4. Questions :**

**Clean the Data**:

Correct the Data missing in Race column

Remove Age column, Url column

Correct the Data missing in Admission column

Make the first row as header & keep the first letter as capital

Correct the Data in Major column

**DERIVING RESULTS:**

Question1.Show the Admission status?

Question2: Show the Admission status with respect to Gender?

Question3: Show the Admission status with respect to Race?

Question4: Show the Admission status with respect to the Work Industry?

Question5: Top 15 applicants list with best GMAT score

**Visualization should be Done vis a Dashboard with a slicer as Work Industry**